J. Phys. Chem. A **2002,** *106,* 10707−10717

Charge Model 3: A Class IV Charge Model Based on Hybrid Density Functional Theory with Variable Exchange

Paul Winget, Jason D. Thompson, James D. Xidos, Christopher J. Cramer,* and Donald G. Truhlar*

Department of Chemistry and Supercomputer Institute, 207 Pleasant St. SE, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received: August 20, 2002

We present a new class IV charge model. The model, called Charge Model 3 (CM3), is designed to be able to obtain accurate partial charges from hybrid density functional calculations with a variable amount of Hartree−Fock exchange and with or without diffuse functions in the basis. The model maps atomic partial charges obtained by Löwdin or redistributed Löwdin population analysis into improved (class IV) charges that reproduce accurate charge-dependent observables for molecules containing H, Li, C, N, O, F, Si, S, P, Cl, and Br. The hybrid density functional theory we use here is based on Adamo and Barone's modified Perdew−Wang (mPW) gradient-corrected exchange functional and the PW91 gradient corrected correlation functional. These parametrizations can be used with any arbitrary fraction of Hartree−Fock exchange in conjunction with any of the five basis sets, MIDI!, MIDI!6D, 6-31G*, 6-31+G*, and 6-31+G**. We also present two parametrizations for Hartree−Fock theory employing the MIDI!6D and 6-31G* basis sets. Dipole moments computed from CM3 atomic point charges have root-mean-square errors between 0.26 and 0.40 D and mean unsigned errors in the range 0.19−0.28 D compared to experiment.

1. Introduction

Partial atomic charges are one of the oldest and most useful concepts in chemistry. Partial charges may be used to explain many qualitative trends in structure and reactivity, and this usage is now complemented by a large number of semiquantitative and quantitative applications, including the use of partial atomic charges in structure−activity correlations, for calculating electrostatic potentials, and for solvation and phase-transfer phenomena. Nevertheless, many quantum mechanics books do not even mention partial atomic charges because the concept is nonrigorous in the absence of a unique definition.

Early attempts to assign partial charges to atoms in molecules used nonquantum mechanical methods such as classical models of dipoles.[1] We have labeled[2] such partial charges class I. Partial charges based on quantum mechanics, but in an arbitrary way, e.g., population analysis[3−7] or assignment of a region of space to an individual atom,[8] are called[2] class II charges. Using a quantum-mechanical wave function to predict a physical observable, such as the electrostatic potential[7,9−20] or dipole moment,[21] and finding charges that best reproduce this observable yields class III charges. The algorithms that yield such charges are somewhat unstable for interior (buried) atoms in molecules,[15] and furthermore, these charges are highly dependent on the quality of the wave function from which they are derived. Class IV charges have been developed to start with wave function-dependent charges and empirically map them to reproduce *experimental* or converged theoretical charge-dependent observables.

Our previous class IV partial charge models[2,22−24] were based on semiempirical molecular orbital theory,[25] ab initio Hartree−Fock theory[26] (HF), density-functional theory[27] (DFT), and hybrid Hartree−Fock density functional theory[28] (hereafter called hybrid DFT). The Hartree−Fock orbitals are eigenfunctions of the Fock operator, which accounts in a mean-field way for Coulomb and exchange interactions but not dynamical electron correlation. DFT orbitals, by contrast, are eigenfunctions of the Kohn−Sham operator, which minimizes a functional incorporating statistical exchange and the energetic effect of electron correlation. The hybrid DFT orbitals are eigenfunctions of an operator (called the Fock-Kohn−Sham operator) that involves a linear combination of terms from the Fock and Kohn−Sham operators. It has become very clear from numerous studies that hybrid DFT can achieve very high accuracy, but the fraction of Hartree−Fock exchange that should be included in the hybrid operator is not a priori clear.[29,30] In fact, the best results are often obtained if this function is reoptimized for the specific problem at hand.[31a,b]

Several hybrid functionals that have been used in the literature since the original B3PW91 method[32] of Becke include the popular B3LYP[33] and BH&HLYP methods. Although these methods have provided results for many applications, our recent work has shown that the modified version of the Perdew−Wang gradient-corrected exchange functional, mPW,[34] when employed with the Perdew−Wang gradient corrected correlation functional,[35] can provide accurate thermochemical data and activation barriers for various chemical reactions.[31,36] One focus of the current work is to extend class IV charge models to mPW hybrid functionals with variable amounts of HF exchange.

Another focus of the current work is the expansion of the database from which the parameters for class IV charges are determined. In our previous class IV charge model (CM2), the training set for compounds containing H, C, N, O, F, Si, P, S, Cl, and Br consisted of 204 compounds. (There are an additional five I containing compounds in that set that are not used in the current work, which does not include I.) For the present work, the number of data for these elements was expanded to 398 by using an expanded data set developed more recently.[37] The new data set is notably more diverse with respect to the representa-

---

* Corresponding authors contact information: C.J.C.: Phone: 612-624-0859. Fax: 612-626-2006. E-mail: cramer@chem.umn.edu. D.G.T.: Phone: 612-624-7555. Fax: 612-626-9390. E-mail: truhlar@umn.edu.

10.1021/jp021917i CCC: $22.00

tions of amides, heteroatom−heteroatom bonds (especially N−O, S−N, S−O, O−Si, F−Si, and Cl−Si bonds), nitrogen-containing heterocycles, Si−H bonds, P−H bonds, and on halogenated compounds. We also added an additional 16 data for compounds that contain lithium, so the current charge model is now applicable to that element. The new data set has a total of 398 data for polar molecules containing H, Li, C, N, O, F, Si, P, S, Cl, and Br.

2. Theory

Two standard methods for the generation of class II charges from wave functions are due to Mulliken[3−5] and Löwdin.[6] In the development of the CM2 model, these two methods were compared.[22] The errors in the dipole moments calculated from the Löwdin charges were about two to four times smaller than those for Mulliken charges, and this has been confirmed more recently[37] for our current data set. In addition, Mulliken overlap populations sometimes take on nonphysical values; Löwdin populations are not so ill behaved. More recently, a redistributed Löwdin analysis has been shown to yield more stable results than either previous method for basis sets containing diffuse functions.[37] Because of this, we chose Löwdin population analysis (LPA) as the starting point for our mapping scheme for nondiffuse basis sets and redistributed Löwdin population analysis (RLPA) as the starting point for diffuse basis sets.

The motivation for using a map to produce a class IV charge model is that the errors in class II charges tend to be systematic for a given kind of bond. By parametrizing the model to reduce the error in the overall dipole moment, these systematic errors in the individual bond dipoles can be reduced.

The function that we use for mapping the class II charges is the same as for the CM2 model. The mapped (CM3) charge is a function of the LPA or RLPA charge, q_k^0, and contains a term, $T_{kk'}(B_{kk'})$, that is a function of the Mayer bond order,[38−40] $B_{kk'}$, between two atoms k and k':

$$q_k = q_k^0 + \sum_{k' \neq k} T_{kk'}(B_{kk'}) \tag{1}$$

This $T_{kk'}$ term is a quadratic function of the Mayer bond order:

$$T_{kk'}(B_{kk'}) = (D_{Z_k Z_{k'}} + C_{Z_k Z_{k'}} B_{kk'}) B_{kk'} \tag{2}$$

where Z_k is the atomic number of atom k and $C_{ZZ'}$ and $D_{ZZ'}$ are the CM3 parameters. Note that we will often substitute the atomic symbol for the atomic number, e.g., D_{67} will be written D_{CN}. Because the total charge remains constant, the charge transferred from k' to k should be equal but opposite to the charge transferred from k to k'. The conservation of charge is maintained by the following relations:

$$C_{ZZ'} = -C_{Z'Z} \tag{3}$$

$$D_{ZZ'} = -D_{Z'Z} \tag{4}$$

2.1. MPW*X*. Work carried out using hybrid Hartree−Fock density functional theory has shown that different amounts of the Hartree−Fock (HF) nonlocal exchange operator are optimum for varying applications. Thus, it is useful to consider a one-parameter set of variable-exchange hybrid Fock−Kohn−Sham operators that are written as follows:

$$F = F^{\mathrm{H}} + \frac{X}{100}F^{\mathrm{HFE}} + \left(1 - \frac{X}{100}\right)(F^{\mathrm{SE}} + F^{\mathrm{GCE}}) + F^{\mathrm{C}} \tag{5}$$

where F^{H} is the Hartree operator, F^{HFE} is the Hartree−Fock exchange operator, X is the percentage of Hartree−Fock exchange, F^{SE} is Slater's local density functional for exchange, F^{GCE} is the gradient correction for the exchange functional, and F^{C} is the total correlation functional including both local and gradient-corrected parts. Several possibilities exist for the exchange and correlation functional; however, we will restrict this discussion to the modified version of the Perdew−Wang gradient-corrected exchange functional (mPW) based on the work of Adamo and Barone[34] and the Perdew−Wang 1991 gradient-corrected correlation functional[35] (PW91). The choice of a numerical value for X is not straightforward. Previous models have set $X = 25$, as in the mPW1PW91[34] method, or $X = 42.8$, as in the MPW1K[31a] method, or $X = 40.6$, as in the *m*PW1N method.[31b] The amount of Hartree−Fock exchange in the MPW1K method was determined by optimization over a series of 40 barrier heights and 20 energies of reaction. By using a different data set, e.g., barriers for nucleophilic substitution reactions, conformational energies, or bond energies in bromine-containing compounds, it is possible that X will take on a different value. Because X can take on any value from 0 to 100, methods of this type can be defined as MPW*X* methods (where X is the percentage Hartree−Fock exchange). (Thus MPW1K is MPW42.8, *m*PW1N is MPW40.6, and mPW1PW91 is MPW25) The goals for this work are (i) to parametrize class IV charge models for MPW*X* using a variety of basis sets and (ii) to find a functional form that describes these parameters as a function of X.

To determine a function for an arbitrary percentage of HF exchange, we will parametrize at five different values of X. This includes three percentages that have been previously used, 0, 25, and 42.8, corresponding to mPWPW91,[34] mPW1PW91,[34] and MPW1K.[31] The two additional points were chosen to cover the other extreme, 99.9, and another point in the chemically interesting region, 60.6. The value of 60.6, which may be denoted as MPW60.6 (or MPW1KK, which is a special name for this case), gives three equally spaced points (25, 42.8, and 60.6).

2.2. Basis Sets. MIDI!-type basis sets[41−43] are used in both the 5D (spherical harmonic) and 6D (Cartesian) form. Note that MIDI! itself is a 5D basis set and is synonymous with MIDIX, MIDI!5D, MIDI!(5D), and MIDIX5D. The version with Cartesian *d* functions is analogously called MIDI!6D, MIDI!(6D), or MIDIX6D; we will call it MIDI!6D. The 6-31G*,[44−46] 6-31+G*,[47] and 6-31+G**[48] basis sets are defined to use the Cartesian *d* sets. Because the 6-31G* and 6-31+G* basis sets are not defined uniquely for bromine, the Binning−Curtiss[49] basis sets, respectively unaugmented and augmented with diffuse functions (diffuse *s* and *p* functions, both with the exponent 0.035), were used in their stead for Br.

3. Parametrization

3.1. Training Sets. The training set used in this work has been taken from ref 37. Becasue amides represent such an important class of chemicals, it is important to note that special efforts were made to have them well represented in the data set of ref 37. It is possible for amides to be either planar or pyramidal with respect to the C−N(R)−R′ improper dihedral angle. Because experimental values of dipole moments[50] are vibrationally averaged and because the potential energy function for amides is usually quite flat with respect to this degree of freedom and because the dipole moment can vary fairly significantly, typically approximately 0.2 D with respect to this angle (see Table 1), fitting single-geometry dipole moments to experiment is hard to justify for amides. Thus, it is significant

TABLE 1: Differences in mPW1PW91/MG3S//HF/MIDI! Density Dipole Moments (debyes)[a] for Different Conformations of Amides

compound	planar	pyramidal	expt
ethanamide (acetamide)	3.91	3.64	3.68[b]
E−*N*-methylacetamide	4.15	3.99	
Z−*N*-methylacetamide	3.87	3.67	
N,*N*-dimethyl-acetamide	3.77	3.96	
formamide	3.96	3.68	3.73[b]
N,*N*-dimethylformamide	4.17	4.11	
E−*N*-methylformamide	4.29	4.14	3.83[b]

[a] Density dipole moments are defined in Section 3.6. [b] Reference 50.

TABLE 2: Bond Orders between Selected Atom Pairs Using mPW1PW91

molecule	atom pair	MIDI!	MIDI!6D	6-31G*	6-31+G*	6-31+G**
ethene	C−H	0.944	0.944	0.939	0.920	0.928
benzene	C−H	0.939	0.939	0.926	0.788	0.770
	differ.	0.005	0.005	0.014	0.132	0.159
methanol	O−H	0.824	0.825	0.791	0.739	0.850
ethanol	O−H	0.821	0.822	0.788	0.724	0.841
	differ.	0.003	0.003	0.003	0.015	0.009
methanol	C−O	0.894	0.891	0.964	0.861	0.843
ethanol	C−O	0.876	0.870	0.950	0.883	0.928
	differ.	0.018	0.021	0.014	0.022	0.086
methyl ether	C−O	0.865	0.858	0.937	0.825	0.810
ethyl ether	C−O	0.848	0.839	0.919	0.873	0.874
	differ.	0.017	0.019	0.018	0.028	0.064
formaldehyde	C=O	1.867	1.846	1.998	1.898	1.888
acetaldehyde	C=O	1.851	1.825	1.969	1.944	1.928
	differ.	0.016	0.021	0.029	0.046	0.040
acetone	C=O	1.830	1.800	1.941	2.069	2.057
2-butanone	C=O	1.825	1.794	1.921	1.986	1.975
	differ.	0.006	0.006	0.020	0.082	0.082
methylamine	C−N	0.939	0.944	0.996	0.926	0.927
ethylamine	C−N	0.947	0.952	1.009	0.953	0.931
	differ.	0.008	0.008	0.013	0.027	0.004
fluroromethane	C−F	0.841	0.830	0.919	0.783	0.771
fluroroethane	C−F	0.016	0.019	0.017	0.034	0.039
	differ.	0.016	0.019	0.017	0.034	0.039
chloromethane	C−Cl	0.966	0.967	0.983	0.950	0.975
chloroethane	C−Cl	0.969	0.969	0.978	0.954	0.984
	differ.	0.003	0.002	0.005	0.004	0.009

TABLE 3: Parameters, Löwdin Charges, and Mapped Charges Using C_{HC} and D_{HC} for Ethene and Benzene Using mPW1PW91

			ethene			benzene		
basis set	C_{HC}	D_{HC}	q_H^0 [a,b]	q_H using C_{HC} [c]	q_H using D_{HC} [d]	q_H^0	q_H using C_{HC}	q_H using D_{HC}
MIDI!	0.003	0.003	0.106	0.108	0.108	0.109	0.112	0.112
MIDI!6D	0.003	0.003	0.106	0.108	0.108	0.109	0.112	0.112
6-31G*	−0.065	−0.061	0.167	0.109	0.110	0.167	0.111	0.110
6-31+G*	−0.090	−0.073	0.179	0.103	0.112	0.174	0.117	0.108
6-31+G**	−0.025	−0.019	0.130	0.109	0.112	0.126	0.111	0.108

[a] The Löwdin charge on H. [b] When the 6-31+G* and 6-31+G** basis sets are used, RLPA charges are used instead of Löwdin charges. [c] The CM3 charge on H using only C_{HC}. [d] The CM3 charge on H using only D_{HC}.

TABLE 4: Root-Mean Square (RMS) Errors (in debyes) for Different Combinations of $C_{ZZ'}$ Parameters Using mPW1PW91/MIDI! for Compounds Containing H, C, N, and O

type of compound	no.[a]	none	C_{CN}	C_{CO}	C_{NO}	all 3
inorganic compounds	6	0.29	0.31	0.29	0.33	0.31
alcohols, phenol	13	0.27	0.26	0.14	0.26	0.14
ethers	11	0.53	0.52	0.23	0.51	0.23
aldehydes	5	0.10	0.10	0.22	0.11	0.19
ketones	11	0.30	0.31	0.30	0.32	0.30
carboxylic acids	9	0.27	0.26	0.26	0.26	0.24
esters	6	0.27	0.27	0.12	0.27	0.11
other C, H, O compounds	12	0.37	0.37	0.28	0.36	0.29
aliphatic amines, aniline	13	0.35	0.28	0.34	0.34	0.28
aromatic nitrogen heterocycles	11	0.25	0.34	0.26	0.25	0.32
nitriles	12	0.20	0.18	0.20	0.20	0.19
imines	6	0.50	0.42	0.51	0.50	0.43
other C, H, N compounds	14	0.19	0.13	0.18	0.19	0.12
amides and phenylurea	17	0.28	0.26	0.15	0.22	0.11
nitrohydrocarbons	5	0.18	0.17	0.19	0.19	0.18
bifunctional H, C, N, and O compounds	12	0.21	0.23	0.30	0.24	0.32
subtotal	163	0.30	0.29	0.25	0.30	0.24

[a] Number of data in the training set for this row.

TABLE 5: Root-Mean Square (RMS) Errors (in debyes) for Different Combinations of $C_{ZZ'}$ Parameters Using mPW1PW91/MIDI! for Sulfur-Containing Compounds

type of compound	no.[a]	none	C_{HS}	C_{CS}	C_{OS}	all 3
inorganic compounds	2	0.07	0.06	0.06	0.05	0.04
thiols	8	0.11	0.15	0.12	0.11	0.16
sulfides, disulfides	9	0.15	0.16	0.17	0.16	0.18
other sulfur-containing compounds	23	0.48	0.46	0.47	0.47	0.45
subtotal	42	0.36	0.36	0.36	0.36	0.35

[a] Number of data in the training set for this row.

that, in addition to the three pyramidal amides that we used previously, the new test set contains the three corresponding planar conformers of formamide, acetamide, and *N*-methylformamide and the pyramidal and planar conformations of five new amides. For each conformer of each amide, we used theoretical dipole moments.

The training set used in this paper also includes the nonpolar molecules ethene and benzene, which are used for obtaining C−H parameters as explained below. For ethene and benzene, we used geometries optimized by the mPW1PW91 method with the MG3S basis set. The MG3S basis is the "modified G3 semidiffuse" basis set,[51] which is obtained from the MG3[52] basis set (which is synonymous with the G3LargeMP2[53] basis set) by deleting diffuse functions on H. This basis is used with spherical harmonic (5D and 7F) polarization functions.

3.2. Software. All calculations were carried out using Gaussian 98[54] with mn-gsm − version 1.8.[55] The error in Gaussian 98 for mPW hybrid functional has been corrected in our version of Gaussian 98 as explained elsewhere.[51]

3.3. Nonlinear Optimization. The scalar dipole moment, μ, is calculated from atomic charges, q_k, by

$$\mu = \sqrt{(\sum_k q_k x_k)^2 + (\sum_k q_k y_k)^2 + (\sum_k q_k z_k)^2} \qquad (6)$$

where x_k, y_k, and z_k are the Cartesian coordinates of atom k. The optimization of $C_{ZZ'}$ and $D_{ZZ'}$ parameters is done through a nonlinear fit using standard NAG Fortran 90 routines,[56] which optimize parameters by taking steps in the Gauss−Newton direction or with the second derivatives taken into account using a quasi-Newton updating scheme. This optimization algorithm employs numerical derivatives with respect to $C_{ZZ'}$ and $D_{ZZ'}$ parameters.

3.4. Dipole Moments. Experimental dipole moments were taken from several compilations[50,57−61] and several additional sources.[62−64] All theoretical dipole moments used for training the model are obtained as one-electron expectation values of the full electron density from single-point mPW1PW91/MG3S calculations. Dipole moments obtained from the full electron

TABLE 6: Root-Mean Square (RMS) Errors (in debyes) for Different Combinations of $C_{ZZ'}$ Parameters Using mPW1PW91/MIDI! for Silicon-Containing Compounds

type of compound	no.[a]	none	C_{CSi}	C_{OSi}	both
C, H, and Si	9	0.09	0.09	n/a	n/a
C, H, O, and Si	9	0.38	0.37	0.25	0.26
C, H, Si, and halogen	18	0.18	0.17	0.19	0.17
subtotal	36	0.23	0.22	0.18	0.18

[a] Number of data in the training set for this row.

TABLE 7: Root-Mean Square (RMS) Errors (in debyes) for Different Combinations of $C_{ZZ'}$ Parameters Using mPW1PW91/MIDI! for Phosphorus-Containing Compounds

type of compound	no.[a]	none	C_{CP}	C_{NP}	C_{OP}	C_{SP}
inorganic compounds	2	0.34	0.32	0.34	0.31	0.28
phosphorus	10	0.19	0.16	0.19	0.16	0.17
multifunctional phosphorus	13	0.29	0.27	0.29	0.18	0.17
compounds with both S and P	7	0.27	0.28	0.27	0.28	0.16
subtotal	32	0.26	0.25	0.26	0.21	0.18

[a] Number of data in the training set for this row

TABLE 8: Parameters, Slopes ($m_{ZZ'}^{[2]}$ and $m_{ZZ'}^{[1]}$), and Intercepts ($b_{ZZ'}$) for Various Percentages of Hartree−Fock Exchange Using the MIDI! Basis Set

	0	25	42.8	60.6	99.9	$m_{ZZ'}^{[2]}$	$m_{ZZ'}^{[1]}$	$b_{ZZ'}$
$C_{ZZ'}$								
C−O	0.066	0.074	0.078	0.082	0.090	−0.0080	0.0310	0.0660
O−Si	−0.061	−0.064	−0.065	−0.066	−0.068	0.0040	−0.0110	−0.0610
O−P	−0.027	−0.025	−0.023	−0.021	−0.018	−0.0030	0.0120	−0.0270
P−S	−0.101	−0.102	−0.103	−0.105	−0.114	−0.0150	0.0020	−0.1010
$D_{ZZ'}$								
H−C	0.008	0.003	−0.001	−0.004	−0.012	0.0000	−0.0203	0.0082
H−N	0.117	0.114	0.112	0.110	0.105	−0.0010	−0.0100	0.1170
H−O	0.034	0.028	0.025	0.021	0.011	−0.0020	−0.0200	0.0340
H−Si	0.032	0.039	0.044	0.049	0.059	0.0000	0.0276	0.0322
H−P	−0.038	−0.037	−0.036	−0.035	−0.032	0.0020	0.0050	−0.0380
H−S	−0.086	−0.083	−0.081	−0.080	−0.076	−0.0010	0.0110	−0.0860
Li−C	0.014	−0.067	−0.127	−0.187	−0.322	0.0000	−0.3369	0.0161
Li−N	0.191	0.168	0.147	0.126	0.081	−0.0130	−0.1000	0.1920
Li−O	0.236	0.183	0.151	0.121	0.058	0.0350	−0.2110	0.2350
Li−F	0.325	0.299	0.284	0.269	0.236	0.0130	−0.1010	0.3240
Li−S	0.061	0.026	0.002	−0.021	−0.070	0.0000	−0.1303	0.0591
Li−Cl	−0.022	−0.055	−0.077	−0.098	−0.146	0.0000	−0.1237	−0.0231
C−N	0.057	0.051	0.046	0.042	0.032	0.0000	−0.0246	0.0567
C−O	−0.073	−0.099	−0.116	−0.133	−0.168	0.0100	−0.1040	−0.0740
C−F	0.028	−0.003	−0.030	−0.054	−0.108	0.0000	−0.1372	0.0292
C−Si	0.079	0.091	0.099	0.108	0.126	0.0000	0.0463	0.0795
C−P	−0.026	−0.024	−0.023	−0.021	−0.018	0.0000	0.0080	−0.0260
C−S	0.017	0.015	0.013	0.011	0.006	−0.0020	−0.0090	0.0170
C−Cl	0.017	0.013	0.010	0.007	0.002	0.0030	−0.0180	0.0170
C−Br	0.078	0.065	0.060	0.056	0.047	0.0190	−0.0470	0.0760
N−O	0.026	0.007	−0.011	−0.025	−0.056	0.0030	−0.0880	0.0280
N−P	−0.115	−0.111	−0.108	−0.105	−0.098	0.0000	0.0168	−0.1152
O−Si	0.021	0.042	0.056	0.070	0.100	0.0000	0.0796	0.0215
O−P	−0.039	−0.030	−0.023	−0.016	0.003	0.0080	0.0340	−0.0390
O−S	−0.028	−0.005	0.012	0.029	0.067	0.0000	0.0959	−0.0289
F−Si	−0.092	−0.061	−0.039	−0.017	0.034	0.0000	0.1265	−0.0926
F−P	−0.095	−0.070	−0.052	−0.034	0.006	0.0000	0.1010	−0.0951
Si−Cl	−0.010	−0.020	−0.027	−0.033	−0.048	0.0000	−0.0384	−0.0099
P−S	0.232	0.227	0.226	0.226	0.232	0.0270	−0.0260	0.2320
P−Cl	0.106	0.101	0.097	0.093	0.086	0.0020	−0.0230	0.1060

TABLE 9: Parameters, Slopes ($m_{ZZ'}^{[2]}$ and $m_{ZZ'}^{[1]}$), and Intercepts ($b_{ZZ'}$) for Various Percentages of Hartree−Fock Exchange Using the 6-31G* Basis Set

	0	25	42.8	60.6	99.9	$m_{ZZ'}^{[2]}$	$m_{ZZ'}^{[1]}$	$b_{ZZ'}$
$C_{ZZ'}$								
C−O	0.056	0.059	0.060	0.061	0.061	−0.0070	0.0120	0.0560
O−Si	−0.082	−0.086	−0.089	−0.091	−0.095	0.0050	−0.0170	−0.0820
O−P	−0.083	−0.082	−0.082	−0.092	−0.095	−0.0090	−0.0100	−0.0800
P−S	0.020	0.012	0.006	−0.063	−0.085	−0.0270	−0.1210	0.0410
$D_{ZZ'}$								
H−C	−0.056	−0.061	−0.065	−0.069	−0.077	0.0000	−0.0210	−0.0561
H−N	0.046	0.041	0.036	0.032	0.022	0.0000	−0.0240	0.0466
H−O	−0.033	−0.040	−0.046	−0.051	−0.063	0.0000	−0.0303	−0.0329
H−Si	−0.002	0.007	0.013	0.019	0.032	0.0000	0.0338	−0.0016
H−P	0.049	0.056	0.061	0.063	0.075	0.0000	0.0240	0.0500
H−S	−0.013	−0.011	−0.010	−0.009	−0.006	0.0000	0.0076	−0.0134
Li−C	0.492	0.491	0.492	0.494	0.501	0.0160	−0.0060	0.4910
Li−N	0.686	0.698	0.709	0.723	0.762	0.0370	0.0400	0.6860
Li−O	0.678	0.694	0.708	0.726	0.775	0.0460	0.0510	0.6780
Li−F	0.595	0.608	0.620	0.635	0.675	0.0370	0.0430	0.5950
Li−S	0.540	0.542	0.544	0.547	0.554	0.0070	0.0070	0.5400
Li−Cl	0.578	0.593	0.604	0.615	0.642	0.0090	0.0550	0.5780
C−N	0.082	0.077	0.074	0.071	0.062	−0.0030	−0.0160	0.0820
C−O	−0.028	−0.045	−0.056	−0.067	−0.088	0.0100	−0.0690	−0.0280
C−F	0.037	0.010	−0.009	−0.027	−0.067	0.0000	−0.1038	0.0360
C−Si	0.089	0.102	0.112	0.121	0.140	0.0000	0.0512	0.0895
C−P	0.110	0.115	0.118	0.125	0.131	−0.0020	0.0260	0.1090
C−S	0.124	0.124	0.123	0.122	0.120	−0.0030	−0.0010	0.1240
C−Cl	0.087	0.082	0.079	0.076	0.069	0.0030	−0.0200	0.0870
C−Br	0.049	0.043	0.038	0.033	0.023	0.0000	−0.0258	0.0490
N−O	0.008	−0.016	−0.030	−0.044	−0.076	0.0100	−0.0920	0.0070
N−P	−0.017	−0.011	−0.006	−0.002	0.009	0.0000	0.0258	−0.0171
O−Si	0.155	0.182	0.200	0.218	0.257	0.0000	0.1020	0.1559
O−P	0.208	0.227	0.242	0.273	0.312	0.0160	0.0990	0.2020
O−S	0.089	0.117	0.138	0.159	0.205	0.0000	0.1169	0.0881
F−Si	0.071	0.111	0.140	0.169	0.235	0.0000	0.1641	0.0699
F−P	0.122	0.156	0.180	0.218	0.274	0.0060	0.1540	0.1170
Si−Cl	−0.018	−0.034	−0.045	−0.056	−0.079	0.0000	−0.0607	−0.0185
P−S	−0.053	−0.049	−0.045	0.045	0.062	0.0310	0.1420	−0.0820
P−Cl	−0.067	−0.078	−0.086	−0.093	−0.109	0.0000	−0.0421	−0.0674

density are called density dipole moments to distinguish them from dipole moments calculated from partial atomic charges.

3.5. Units. All atomic partial charges are given in atomic units, in which the charge on the proton is unity. All dipole moments are in debyes.

3.6. Parametrization Procedure. *3.6.1 Choice of $C_{ZZ'}$ and $D_{ZZ'}$ Parameters.* In our previous class IV charge model, we used both $C_{ZZ'}$ and $D_{ZZ'}$ parameters. Table 2 shows that, for basis sets containing diffuse functions, the bond orders can vary significantly for compounds with the same type of bond, even though the bond orders for these compounds are nearly constant when using basis sets without diffuse functions. For example, using mPW1PW91, the differences in the C−O bond orders between acetone and 2-butanone are 0.08 for 6-31+G* and 6-31+G**, which is over four times greater than the corresponding differences for basis sets that do not have diffuse functions. Because the mapping scheme assumes that chemically similar bonds have similar bond orders, this could lead to decreased accuracy. Because $C_{ZZ'}$ parameters are multiplied by the bond order squared whereas $D_{ZZ'}$ parameters are multiplied by the first power of the bond order, the overall charge model corrections obtained with $C_{ZZ'}$ parameters are more sensitive to these differences in the bond order. This problem becomes more pronounced when solving for the electronic structure in the presence of a reaction field that is created by these atomic partial charges, which introduces an even greater degree of nonlinearity. Therefore, for a given pair of atoms, we decided that our charge model strategy would be to use $D_{ZZ'}$ in preference to $C_{ZZ'}$. We made exceptions to this rule if using a particular $C_{ZZ'}$ parameter exclusively or using both types of parameters yielded significantly more accurate dipole moments than when using only the $D_{ZZ'}$ parameter. We found no examples of the former, but, as explained below, we did sometimes use both types of parameters.

Note that another approach for improving the charge model would be to adopt a different model for the bond order, if one could be found that were less sensitive to the presence of diffuse functions in the basis set. However, previous work has indicated[65,66] that the Mayer definition is less sensitive to basis set extension than other methods. Furthermore, the particularly simple functional dependence of the Mayer bond order on

TABLE 10: Parameters, Slopes ($m_{ZZ'}^{[2]}$ and $m_{ZZ'}^{[1]}$), and Intercepts ($b_{ZZ'}$) for Various Percentages of Hartree−Fock Exchange Using the 6-31+G Basis Set**

	0	25	42.8	60.6	99.9	$m_{ZZ'}^{[2]}$	$m_{ZZ'}^{[1]}$	$b_{ZZ'}$
				$C_{ZZ'}$				
C−O	0.025	0.028	0.030	0.031	0.034	−0.0040	0.0140	0.0250
O−Si	−0.099	−0.100	−0.101	−0.102	−0.099	0.0110	−0.0120	−0.0980
O−P	−0.177	−0.212	−0.223	−0.235	−0.257	0.0550	−0.1280	−0.1800
P−S	−0.024	−0.081	−0.092	−0.103	−0.132	0.0970	−0.1900	−0.0320
				$D_{ZZ'}$				
H−C	−0.017	−0.019	−0.020	−0.021	−0.024	0.0010	−0.0070	−0.0170
H−N	0.153	0.149	0.146	0.142	0.135	0.0000	−0.0172	0.1528
H−O	0.142	0.139	0.137	0.135	0.132	0.0060	−0.0170	0.1420
H−Si	−0.004	0.001	0.005	0.009	0.016	−0.0030	0.0230	−0.0040
H−P	0.062	0.073	0.081	0.087	0.099	−0.0110	0.0480	0.0620
H−S	0.004	0.006	0.008	0.011	0.019	0.0090	0.0070	0.0040
Li−C	0.369	0.355	0.346	0.338	0.322	0.0110	−0.0570	0.3680
Li−N	0.473	0.460	0.451	0.443	0.425	0.0050	−0.0540	0.4730
Li−O	0.448	0.441	0.435	0.430	0.418	0.0000	−0.0302	0.4483
Li−F	0.515	0.521	0.525	0.527	0.532	−0.0090	0.0260	0.5150
Li−S	0.435	0.433	0.431	0.429	0.426	0.0020	−0.0110	0.4350
Li−Cl	0.434	0.429	0.425	0.421	0.412	0.0000	−0.0214	0.4339
C−N	0.085	0.080	0.077	0.073	0.063	−0.0050	−0.0170	0.0850
C−O	−0.001	−0.018	−0.031	−0.044	−0.071	0.0000	−0.0704	−0.0009
C−F	−0.015	−0.050	−0.074	−0.098	−0.149	0.0000	−0.1343	−0.0157
C−Si	−0.002	0.006	0.011	0.015	0.023	−0.0070	0.0320	−0.0020
C−P	0.169	0.170	0.170	0.169	0.170	−0.0010	0.0010	0.1700
C−S	0.152	0.151	0.149	0.148	0.144	−0.0040	−0.0040	0.1520
C−Cl	0.110	0.103	0.099	0.095	0.086	0.0020	−0.0260	0.1100
C−Br	0.104	0.100	0.097	0.093	0.086	0.0000	−0.0185	0.1043
N−O	−0.025	−0.043	−0.054	−0.064	−0.087	0.0110	−0.0720	−0.0250
N−P	0.050	0.056	0.060	0.064	0.073	0.0000	0.0226	0.0503
O−Si	0.270	0.292	0.308	0.322	0.350	−0.0160	0.0970	0.2700
O−P	0.469	0.550	0.585	0.620	0.688	−0.1010	0.3090	0.4740
O−S	0.157	0.191	0.215	0.240	0.293	0.0000	0.1366	0.1566
F−Si	0.176	0.212	0.238	0.264	0.319	0.0000	0.1430	0.1767
F−P	0.339	0.376	0.426	0.479	0.621	0.1470	0.1420	0.3350
Si−Cl	−0.018	−0.029	−0.036	−0.043	−0.056	0.0070	−0.0450	−0.0180
P−S	0.050	0.148	0.163	0.179	0.217	−0.1850	0.3240	0.0640
P−Cl	−0.089	−0.097	−0.103	−0.108	−0.120	0.0000	−0.0313	−0.0892

density matrix elements makes it an ideal choice for use in situations where analytic derivatives of the CM3 charges with respect to density matrix elements are needed, e.g., in self-consistent-reaction-field calculations using the SM*x* family of continuum solvation models.[67−72]

3.6.2. C−H Bond Dipole Moment. One of the limitations of using the molecular dipole moment for parametrization is that for some compounds significantly different charge distributions can give the same dipole moment. For example, in methane, ethene, and benzene, the dipole moment is zero regardless of the charge on C or H, as long as the charge distributions have the correct symmetry. Because C−H bonds occur in nearly every organic compound, it is important to devise a parametrization strategy that gives physically reasonable C−H bond dipoles.

We will optimize a C−H parameter by requiring the average charge on H in ethene and benzene to be 0.11, a value which is justified in our previous work.[22] Table 3 shows the results for this parametrization when using either the C_{HC} or the D_{HC} parameter. When the C_{HC} parameter is used, the difference between the CM3 charge on the hydrogen in benzene and in ethene becomes large when diffuse functions are added. When the D_{HC} parameter is used, the resulting CM3 charge on hydrogen in ethene and in benzene are closer to the value of 0.11, even for 6-31+G* and 6-31+G**. Although the differences observed when using the D_{HC} parameter compared to when using the C_{HC} parameter are small for 6-31+G* and 6-31+G**, the bond order (see Table 2) between C and H in benzene and ethene is sensitive to the inclusion of diffuse basis functions. Therefore, we used the D_{HC} parameter to avoid potential problems in subsequent stages of the parameter optimization.

TABLE 11: Root-Mean Square (RMS) Errors (in debyes) for Various Percentages of Hartree−Fock Exchange Using the MIDI! Basis Set Broken Down by Functional Group

type of compound	no.[a]	0	25	42.8	60.6	99.9
inorganic compounds	10	0.28	0.26	0.25	0.24	0.23
alcohols, phenol	13	0.15	0.14	0.14	0.14	0.14
ethers	11	0.26	0.23	0.21	0.20	0.18
aldehydes	5	0.23	0.22	0.21	0.20	0.19
ketones	11	0.31	0.30	0.29	0.28	0.27
carboxylic acids	9	0.22	0.26	0.28	0.30	0.35
esters	6	0.11	0.12	0.13	0.14	0.15
other C, H, and O compounds	12	0.30	0.28	0.27	0.27	0.26
aliphatic amines, aniline	13	0.34	0.34	0.35	0.35	0.35
aromatic nitrogen heterocycles	11	0.26	0.26	0.26	0.26	0.29
nitriles	12	0.20	0.20	0.20	0.20	0.20
imines	6	0.50	0.51	0.51	0.51	0.51
other C, H, N compounds	14	0.21	0.18	0.17	0.16	0.17
amides and phenylurea	17	0.14	0.15	0.17	0.19	0.23
nitrohydrocarbons	5	0.18	0.19	0.20	0.23	0.32
bifunctional H, C, N, and O compounds	12	0.32	0.30	0.29	0.28	0.26
H, C, N, and O polar compounds[b]	163	0.26	0.25	0.25	0.25	0.26
fluorine-containing compounds	39	0.24	0.22	0.21	0.20	0.20
chlorine-containing compounds	33	0.14	0.15	0.15	0.15	0.15
bromine-containing compounds	14	0.15	0.15	0.14	0.13	0.11
halogenated bifunctional compounds	23	0.32	0.30	0.29	0.28	0.26
thiols	8	0.10	0.11	0.11	0.13	0.16
sulfides, disulfides	9	0.13	0.15	0.16	0.18	0.23
other sulfur-containing compounds	23	0.49	0.48	0.48	0.48	0.50
phosphorus	10	0.13	0.15	0.17	0.19	0.21
multifunctional phosphorus	13	0.17	0.18	0.19	0.20	0.22
compounds with S and P	7	0.09	0.09	0.09	0.09	0.10
C, H, and Si	9	0.08	0.09	0.10	0.10	0.11
C, H, O, and Si	9	0.27	0.25	0.24	0.24	0.22
C, H, Si, and halogen	18	0.19	0.18	0.17	0.16	0.15
lithium compounds	16	0.67	0.63	0.60	0.59	0.55
all polar compounds	398	0.28	0.27	0.27	0.27	0.27

[a] Number of data in the training set for this row. [b] Six of the 10 inorganic compounds contain at most H, C, N, and O.

3.6.3. Other Bond Dipole Moments. For the rest of the parameters, we used three general guidelines to decide whether to use $C_{ZZ'}$ or $D_{ZZ'}$ or both: (i) when choosing the first parameter, the $D_{ZZ'}$ is always used rather than $C_{ZZ'}$, unless the error when using $C_{ZZ'}$was significantly smaller, (ii) if an atom pair only forms one type of bond, e.g., a single bond where $B_{ZZ'} \approx 1$, then only one parameter is needed, and (iii) when added as a second parameter, $C_{ZZ'}$ should significantly decrease the error in the molecules that are affected.

Once the C−H parameter was fixed, we optimized parameters for all of the molecules that contain at most H, C, N, and O. This is a change from previous CM2 parametrizations, where the N−O parameters were optimized with the O−S and O−P parameters. Table 4 shows, for mPW1PW91/MIDI!, the effect of adding additional $C_{ZZ'}$ parameters to the $D_{ZZ'}$ parameters for compounds that contain at most H, C, N, and O. In no case was using $C_{ZZ'}$ better than $D_{ZZ'}$, so the question that remains is whether to use $C_{ZZ'}$ in addition to $D_{ZZ'}$. Upon addition of a C_{CN} parameter, there is improvement over the 13 amines of 0.07 D in the root-mean square (RMS) error; however, this is balanced out by an increase in the RMS error for the aromatic nitrogen heterocycles of 0.09 D. The addition of C_{CO} causes the RMS error for most classes of compounds to decrease, with marked improvements of 0.13 D in the alcohols, 0.3 D in ethers, and 0.15 D in esters. The aldehydes become only slightly less accurate. There is almost no effect on the error by adding C_{NO}, suggesting that the D_{NO} term is able to capture the systematic trends. Adding all three $C_{CZ'}$ parameters causes the error to decrease only slightly from the case where only C_{CO} is added. Adding $C_{ZZ'}$ terms involving hydrogen also had little effect. The same pattern was observed for the other fractions of HF exchange, as well as for all of the MPW*X*/6-31G* levels of

TABLE 12: Root-Mean Square (RMS) Errors (in debyes) for Various Percentages of Hartree−Fock Exchange Using the 6-31G* Basis Set Broken Down by Functional Group

type of compound	no.[a]	0	25	42.8	60.6	99.9
inorganic compounds	10	0.32	0.31	0.31	0.31	0.33
alcohols, phenol	13	0.15	0.15	0.15	0.15	0.15
ethers	11	0.18	0.16	0.15	0.14	0.13
aldehydes	5	0.25	0.22	0.21	0.20	0.19
ketones	11	0.25	0.23	0.21	0.20	0.19
carboxylic acids	9	0.19	0.22	0.24	0.26	0.30
esters	6	0.23	0.22	0.23	0.23	0.24
other C, H, and O compounds	12	0.29	0.27	0.25	0.24	0.23
aliphatic amines, aniline	13	0.25	0.25	0.25	0.24	0.23
aromatic nitrogen heterocycles	11	0.27	0.25	0.25	0.25	0.27
nitriles	12	0.23	0.23	0.23	0.23	0.23
imines	6	0.40	0.40	0.40	0.40	0.40
other C, H, and N compounds	14	0.19	0.16	0.15	0.15	0.16
amides and phenylurea	17	0.14	0.16	0.17	0.19	0.23
nitrohydrocarbons	5	0.15	0.17	0.19	0.22	0.30
bifunctional H, C, N, and O compounds	12	0.29	0.27	0.26	0.25	0.26
H, C, N, and O polar compounds[b]	163	0.23	0.23	0.22	0.22	0.23
fluorine-containing compounds	39	0.22	0.21	0.20	0.20	0.20
chlorine-containing compounds	33	0.15	0.14	0.13	0.13	0.12
bromine-containing compounds	14	0.17	0.15	0.15	0.14	0.12
halogenated bifunctional compounds	23	0.25	0.23	0.22	0.22	0.22
thiols	8	0.12	0.12	0.12	0.14	0.17
sulfides, disulfides	9	0.18	0.20	0.22	0.25	0.31
other sulfur-containing compounds	23	0.52	0.54	0.56	0.59	0.67
phosphorus	10	0.38	0.41	0.43	0.42	0.46
multifunctional phosphorus	13	0.47	0.48	0.49	0.51	0.51
compounds with S and P	7	0.17	0.18	0.19	0.17	0.19
C, H, and Si	9	0.08	0.08	0.09	0.09	0.10
C, H, O, and Si	9	0.26	0.25	0.24	0.23	0.22
C, H, Si, and halogen	18	0.16	0.15	0.14	0.14	0.13
lithium compounds	16	0.30	0.27	0.26	0.26	0.26
all polar compounds	398	0.26	0.26	0.26	0.27	0.28

[a] Number of data in the training set for this row. [b] Six of the 10 inorganic compounds contain at most H, C, N, and O.

TABLE 13: Root-Mean Square (RMS) Errors (in debyes) for Various Percentages of Hartree−Fock Exchange Using the 6-31+G Basis Set Broken Down by Functional Group**

type of compound	no.[a]	0	25	42.8	60.6	99.9
inorganic compounds	10	0.35	0.36	0.37	0.37	0.38
alcohols, phenol	13	0.23	0.24	0.24	0.24	0.25
ethers	11	0.22	0.23	0.23	0.24	0.25
aldehydes	5	0.33	0.30	0.28	0.27	0.26
ketones	11	0.23	0.22	0.21	0.21	0.23
carboxylic acids	9	0.23	0.27	0.29	0.31	0.34
esters	6	0.31	0.27	0.24	0.23	0.20
other C, H, and O compounds	12	0.25	0.25	0.26	0.27	0.28
aliphatic amines, aniline	13	0.27	0.27	0.26	0.26	0.24
aromatic nitrogen heterocycles	11	0.40	0.39	0.38	0.38	0.38
nitriles	12	0.28	0.26	0.26	0.25	0.25
imines	6	0.40	0.40	0.40	0.40	0.40
other C, H, N compounds	14	0.26	0.24	0.23	0.22	0.24
amides and phenylurea	17	0.34	0.36	0.38	0.40	0.44
nitrohydrocarbons	5	0.26	0.24	0.22	0.20	0.18
bifunctional H, C, N, and O compounds	12	0.27	0.27	0.28	0.29	0.33
H, C, N, and O polar compounds[b]	163	0.29	0.28	0.29	0.29	0.30
fluorine-containing compounds	39	0.34	0.38	0.41	0.44	0.50
chlorine-containing compounds	33	0.30	0.27	0.26	0.24	0.22
bromine-containing compounds	14	0.29	0.27	0.26	0.25	0.24
halogenated bifunctional compounds	23	0.31	0.28	0.28	0.27	0.28
thiols	8	0.16	0.15	0.16	0.17	0.21
sulfides, disulfides	9	0.25	0.24	0.23	0.22	0.22
other sulfur-containing compounds	23	0.61	0.63	0.65	0.68	0.76
phosphorus	10	0.31	0.34	0.35	0.35	0.35
multifunctional phosphorus	13	0.27	0.27	0.25	0.24	0.23
compounds with S and P	7	0.26	0.38	0.39	0.41	0.49
C, H, and Si	9	0.22	0.21	0.21	0.21	0.22
C, H, O, and Si	9	0.36	0.36	0.36	0.37	0.38
C, H, Si, and halogen	18	0.62	0.65	0.68	0.72	0.81
lithium compounds	16	0.42	0.41	0.41	0.41	0.42
all polar compounds	398	0.35	0.35	0.36	0.37	0.40

[a] Number of data in the training set for this row. [b] Six of the 10 inorganic compounds contain at most H, C, N, and O.

theory. So, for H, C, N, and O compounds, we chose to use only one $C_{ZZ'}$ parameter, C_{CO}, in addition to the $D_{ZZ'}$ parameters.

Next, with the H, C, N, and O parameters fixed, the parameters for F, Cl, and Br were optimized. The halogens were not improved by the addition of $C_{ZZ'}$ parameters. This is expected because they generally form only single bonds, thus only the $D_{ZZ'}$ term is needed to correct for any systematic deficiencies.

Then, fixing all of the above parameters, we optimized the sulfur parameters. Table 5 shows the RMS errors when adding additional $C_{ZZ'}$ parameters for atom pairs involving sulfur. The RMS error is relatively constant for these compounds regardless of whether one uses any $C_{ZZ'}$ terms. This was surprising because the CM2 model had both C_{CS} and D_{CS} parameters. Nevertheless, we chose to set all $C_{Z\mathrm{S}}$ parameters equal to zero.

Next, we optimized the silicon parameters. We observed the best results when we broke this part of the optimization up into two steps. First, we optimized the D_{HSi} and D_{CSi} parameters and considered adding a C_{CSi} parameter using the subset of the training data set that contains at most C, H, and Si. The inclusion of SiH and SiH_3 are very important in this step in order to get stable results for Si−H bonds, which, like C−H bonds, are almost nonpolar. Then, we held these parameters constant, and we optimized the other silicon parameters by first using the subset of the training set that contains Si and at most C, H, and O, and then using the subset of the training set that contains Si and at most C, H, and *X*, where *X* is F, Cl, or Br. Because there are no compounds in the silicon training set that contain both oxygen and a halogen, these optimization steps can be carried out in two independent steps. We found that the C_{OSi} parameter caused a large decrease in the error of the siloxanes in the data set (Table 6), whereas no other $C_{\mathrm{Si}Z'}$ parameter caused an appreciable increase in accuracy. Therefore, we included only one $C_{\mathrm{Si}Z'}$ parameter.

The last step consisted of fitting the phosphorus and lithium parameters while keeping the parameters from the previous stages fixed. Table 7 lists four different $C_{\mathrm{P}Z'}$ parameters and the effect that adding each of these parameters has on phosphorus-containing compounds. Of the possible $C_{\mathrm{P}Z'}$ parameters, both C_{SP} and C_{OP} have a significant effect on the RMS error.

The final parametrization included only four nonzero $C_{ZZ'}$, all of which significantly increased the accuracy of the affected classes of compounds, and it includes 30 $D_{ZZ'}$ parameters. If all 34 of the parameters are simultaneously relaxed using the stepwise optimized parameters as guesses, then the RMS error decreases by less than 0.01 D from the stepwise optimized parameter's RMS error; therefore, we used the values from the stepwise procedure described above, because the stepwise procedure may give slightly more physical results.

4. Results and Discussion

4.1. Parameters. The CM3 parameters for the five basis sets at five different fractions of Hartree−Fock exchange with the MIDI!, 6-31G*, and 6-31+G** basis sets are listed in Tables 8−10. The parameters for the MIDI!6D and 6-31+G* basis sets follow similar trends and are given in Tables S3 and S4 of the Supporting Information. The absolute value of most of the parameters is less than 0.1. This means that for a typical single bond, which has a bond order approximately equal to 1, the Löwdin charge is usually altered by less than about 0.1. This is encouraging since the mapping procedure will probably work best when the change is a small perturbation.

TABLE 14: Root-Mean Square (RMS) Errors (in debyes) for 25% (mPW1PW91) and 60.6% (MPW1KK) When Using the MIDI! Basis Set for the Dipole Moment When Calculated Using the Density, Löwdin Atomic Partial Charges, and CM3 Partial Charges

		25%			60.6%		
type of compound	no.[a]	density	Löwdin	CM3	density	Löwdin	CM3
inorganic compounds	10	0.29	0.61	0.26	0.31	0.56	0.24
alcohols, phenol	13	0.14	0.21	0.14	0.20	0.33	0.14
ethers	11	0.15	0.43	0.23	0.17	0.61	0.20
aldehydes	5	0.29	0.19	0.22	0.07	0.14	0.20
ketones	11	0.53	0.44	0.30	0.30	0.23	0.28
carboxylic acids	9	0.23	0.25	0.26	0.30	0.37	0.30
esters	6	0.18	0.30	0.12	0.15	0.27	0.14
other C, H, and O compounds	12	0.42	0.34	0.28	0.30	0.45	0.27
aliphatic amines, aniline	13	0.23	0.25	0.34	0.28	0.19	0.35
aromatic nitrogen heterocycles	11	0.24	0.77	0.26	0.16	0.68	0.26
nitriles	12	0.29	0.99	0.20	0.18	0.81	0.20
imines	6	0.30	0.52	0.51	0.31	0.45	0.51
other C, H, N compounds	14	0.27	0.95	0.18	0.17	0.78	0.16
amides and phenylurea	17	0.46	0.47	0.15	0.26	0.28	0.19
nitrohydrocarbons	5	0.37	0.22	0.19	0.16	0.34	0.23
bifunctional H, C, N, and O compounds	12	0.35	0.42	0.30	0.17	0.23	0.28
H, C, N, and O polar compounds[b]	163	0.32	0.55	0.25	0.23	0.48	0.25
fluorine-containing compounds	39	0.44	0.22	0.22	0.20	0.44	0.20
chlorine-containing compounds	33	0.44	0.23	0.15	0.52	0.17	0.15
bromine-containing compounds	14	0.13	0.67	0.15	0.08	0.54	0.13
halogenated bifunctional compounds	23	0.43	0.44	0.30	0.37	0.52	0.28
thiols	8	0.40	0.13	0.11	0.41	0.15	0.13
sulfides, disulfides	9	0.41	0.37	0.15	0.41	0.35	0.18
other sulfur-containing compounds	23	0.43	0.63	0.48	0.42	0.65	0.48
phosphorus	10	0.21	0.30	0.15	0.17	0.26	0.19
multifunctional phosphorus	13	0.39	1.04	0.18	0.21	0.84	0.20
compounds with S and P	7	0.46	0.33	0.09	0.36	0.41	0.09
C, H, and Si	9	0.14	0.50	0.09	0.15	0.52	0.10
C, H, O, and Si	9	0.51	0.95	0.25	0.34	0.76	0.24
C, H, Si, and halogen	18	0.44	0.73	0.18	0.37	0.51	0.16
lithium compounds	16	0.58	0.74	0.63	0.59	0.74	0.59
all polar compounds	398	0.38	0.55	0.27	0.32	0.50	0.27

[a] Number of data in the training set for this row. [b] Six of the 10 inorganic compounds contain at most H, C, N, and O.

TABLE 15: Root-Mean Square (RMS) Errors (in debyes) for 25% (mPW1PW91) and 60.6% (MPW1KK) When Using the 6-31+G Basis Set for the Dipole Moment When Calculated Using the Density, RLPA Atomic Partial Charges, and CM3 Partial Charges**

		25%			60.6%		
type of compound	no.[a]	density	RLPA	CM3	density	RLPA	CM3
inorganic compounds	10	0.33	0.56	0.36	0.38	0.55	0.37
alcohols, phenol	13	0.32	0.24	0.24	0.42	0.22	0.24
ethers	11	0.21	0.25	0.23	0.32	0.38	0.24
aldehydes	5	0.41	0.23	0.30	0.60	0.19	0.27
ketones	11	0.30	0.34	0.22	0.49	0.17	0.21
carboxylic acids	9	0.43	0.40	0.27	0.58	0.46	0.31
esters	6	0.18	0.24	0.27	0.21	0.23	0.23
other C, H, and O compounds	12	0.41	0.33	0.25	0.59	0.36	0.27
aliphatic amines, aniline	13	0.33	0.39	0.27	0.38	0.33	0.26
aromatic nitrogen heterocycles	11	0.15	0.98	0.39	0.20	0.91	0.38
nitriles	12	0.32	1.28	0.26	0.42	1.16	0.25
imines	6	0.39	0.71	0.40	0.46	0.62	0.40
other C, H, and N compounds	14	0.28	1.06	0.24	0.39	0.94	0.22
amides and phenylurea	17	0.09	0.67	0.36	0.23	0.52	0.40
nitrohydrocarbons	5	0.37	0.34	0.24	0.62	0.59	0.20
bifunctional H, C, N, and O compounds	12	0.18	0.49	0.27	0.33	0.42	0.29
H, C, N, and O polar compounds[b]	163	0.30	0.66	0.28	0.42	0.60	0.29
fluorine-containing compounds	39	0.18	0.51	0.38	0.31	0.73	0.44
chlorine-containing compounds	33	0.25	0.92	0.27	0.35	0.84	0.24
bromine-containing compounds	14	0.24	0.99	0.27	0.34	0.91	0.25
halogenated bifunctional compounds	23	0.23	1.05	0.28	0.31	1.16	0.27
thiols	8	0.22	0.93	0.15	0.28	0.94	0.17
sulfides, disulfides	9	0.26	1.50	0.24	0.31	1.50	0.22
other sulfur-containing compounds	23	0.40	1.23	0.63	0.55	1.32	0.68
phosphorus	10	0.13	0.65	0.34	0.14	0.62	0.35
multifunctional phosphorus	13	0.17	1.70	0.27	0.31	1.92	0.24
compounds with S and P	7	0.14	1.68	0.38	0.29	1.85	0.41
C, H, and Si	9	0.09	0.22	0.21	0.10	0.21	0.21
C, H, O, and Si	9	0.18	0.97	0.36	0.30	1.15	0.37
C, H, Si, and halogen	18	0.24	1.27	0.65	0.34	1.40	0.72
lithium compounds	16	0.10	3.50	0.41	0.20	3.32	0.41
all polar compounds	398	0.26	1.13	0.35	0.37	1.14	0.37

[a] Number of data in the training set for this row. [b] Six of the 10 inorganic compounds contain at most H, C, N, and O.

There are several cases where the parameters are larger than 0.1, most notably, the parameters that involve Li for methods using the 6-31G-type basis sets. For MIDI!-type basis sets, the parameters for Li are in the same range as for the other atom pairs. This is a reflection of the fact that the dipole moment calculated from Löwdin charges using the 6-31G-type basis sets are significantly in error. The MIDI! basis set for Li was designed[43] specifically to improve the charges that were determined from Löwdin analysis, so the CM3 corrections to the Löwdin charges is much smaller.

The other cases where the parameters are greater than 0.1 are when there are nonzero values for both the $C_{ZZ'}$ and the $D_{ZZ'}$ parameters. For example, using the 6-31G* basis set, to differentiate between single and double O−P bonds, the value of D_{OP} is large and positive, between 0.208 and 0.312. The value of C_{OP} is negative, which lessens the amount of charge that is transferred in the mapping procedure.

4.2. General Performance. Tables 11, 12, and S4 show that the accuracies of the predicted dipole moments using MIDI!, MIDI!6D, and 6-31G* are relatively similar with an overall RMS error of 0.3 D. Note that there is little difference in the RMS errors between the 5D and the 6D option for MIDI!. Tables 13 and S5 show that using basis sets that have diffuse functions results in larger RMS errors than when using MIDI!, MIDI!6D, or 6-31G*. The accuracy of the mapping is consistent for all five percentages of Hartree−Fock exchange examined, and with the exception of MIDI! and MIDI!6D, the errors when using 99.9% exchange are slightly higher than when using the other four percentages.

Table 14 lists the RMS errors over the entire training set and over various subsets of the training set for dipole moments calculated using the density, the Löwdin atomic partial charges, and the CM3 atomic partial charges for mPW1PW91 and MPW60.6 when using the MIDI! basis set. The RMS error over the entire training set for the CM3 dipole moments is 0.1 D smaller than the RMS error for the density dipole moments and 0.3 D smaller than the RMS error for the Löwdin dipole moments. Table 14 also illustrates that, for a given subset of the training set (i.e., for a given functional group), the RMS error for the CM3 dipole moments remains relatively constant for a wide range of percentages of HF exchange, whereas the RMS errors for both the density dipoles and the Löwdin dipoles vary considerably.

Table 15 lists the RMS errors over the entire training set and over various subsets of the training set for dipole moments calculated using the density, the RLPA atomic partial charges,

TABLE 16: Predicted Parameters Using Linear Regression and Parameters Determined by Parameterization Using MIDI! and 51.7% Hartree−Fock Exchange

	linear regression	direct parametrizn		linear regression	direct parametrizn
			$C_{ZZ'}$		
C−O	0.080	0.080	O−P	−0.022	−0.022
O−Si	−0.066	−0.066	P−S	−0.104	−0.104
			$D_{ZZ'}$		
H−C	−0.002	−0.002	C−Si	0.103	0.104
H−N	0.112	0.111	C−P	−0.022	−0.022
H−O	0.023	0.023	C−S	0.012	0.012
H−Si	0.046	0.047	C−Cl	0.008	0.008
H−P	−0.035	−0.035	C−Br	0.057	0.058
H−S	−0.081	−0.081	N−O	−0.017	−0.018
Li−C	−0.158	−0.157	N−P	−0.107	−0.107
Li−N	0.137	0.136	O−Si	0.063	0.063
Li−O	0.135	0.136	O−P	−0.019	−0.019
Li−F	0.275	0.276	O−S	0.021	0.020
Li−S	−0.008	−0.010	F−Si	−0.027	−0.028
Li−Cl	−0.087	−0.088	F−P	−0.043	−0.043
C−N	0.044	0.044	Si−Cl	−0.030	−0.030
C−O	−0.125	−0.124	P−S	0.226	0.225
C−F	−0.042	−0.042	P−Cl	0.095	0.095

TABLE 17: Root-Mean Square (RMS) Errors for 51.7% HF Exchange with the MIDI! Basis Set Broken Down by Functional Group Using Parameters that Are Predicted from Linear Regression and Those from Direct Parameterization

type of compound	no.[a]	linear regression	direct parametrizn
inorganic compounds	10	0.24	0.24
alcohols, phenol	13	0.14	0.14
ethers	11	0.20	0.21
aldehydes	5	0.20	0.21
ketones	11	0.29	0.28
carboxylic acids	9	0.29	0.29
esters	6	0.13	0.13
other C, H, and O compounds	12	0.27	0.27
aliphatic amines, aniline	13	0.35	0.35
aromatic nitrogen heterocycles	11	0.26	0.26
nitriles	12	0.20	0.20
imines	6	0.51	0.51
other C, H, and N compounds	14	0.17	0.17
amides and phenylurea	17	0.18	0.18
nitrohydrocarbons	5	0.22	0.22
bifunctional H, C, N, and O compounds	12	0.28	0.28
H, C, N, and O polar compounds[b]	163	0.25	0.25
fluorine-containing compounds	39	0.21	0.21
chlorine-containing compounds	33	0.15	0.15
bromine-containing compounds	14	0.14	0.13
halogenated bifunctional compounds	23	0.28	0.28
thiols	8	0.12	0.12
sulfides, disulfides	9	0.17	0.17
other sulfur-containing compounds	23	0.48	0.48
phosphorus	10	0.18	0.18
multifunctional phosphorus	13	0.19	0.19
compounds with S and P	7	0.09	0.09
C, H, and Si	9	0.10	0.10
C, H, O, and Si	9	0.24	0.24
C, H, Si, and halogen	18	0.17	0.17
lithium compounds	16	0.59	0.59
all polar compounds	398	0.27	0.27

[a] Number of data in the training set for this row. [b] Six of the 10 inorganic compounds contain at most H, C, N, and O.

and the CM3 atomic partial charges for mPW1PW91 and MPW60.6 when using the 6-31+G** basis set. When using mPW1PW91, the density dipole moments are slightly more accurate than the CM3 dipole moments, but when the percentage of HF exchange is increased, the errors in the CM3 dipole moments remain constant, whereas the errors in the density dipoles become larger. The RMS error over compounds in the training set that contain at most H, C, N, and O for the RLPA dipole moments are between 0.3 and 0.5 D larger than the RMS error for the CM3 dipole moments. The RMS error over all 398 dipole moments indicates that the CM3 dipole moments are more accurate than the RLPA dipole moments by 0.78 D when using mPW1PW91 and 0.77 D when using MPW1KK.

TABLE 18: Parameters for Hartree−Fock Using MIDI!6D and 6-31G* for Both CM2 and CM3

	HF/MIDI!6D		HF/6-31G*	
	CM2	CM3	CM2	CM3
		$C_{ZZ'}$		
H−C	−0.030		−0.070	
H−N	0.075		0.012	
H−O	−0.002		−0.052	
H−Si	0.144		−0.019	
H−S	−0.042		−0.012	
C−N	0.017		0.010	
C−O	0.084	0.103	0.053	0.064
C−Si	0.026		0.109	
C−S	−0.033		−0.079	
N−O	0.013		0.000	
O−Si		0.034		−0.097
O−P		−0.072		−0.075
P−S		−0.027		−0.025
		$D_{ZZ'}$		
H−C		0.008		−0.056
H−N		0.099		0.040
H−O		−0.012		−0.046
H−Si		0.045		0.043
H−P	−0.036	−0.055	0.069	0.111
H−S		−0.092		−0.234
Li−C	*a*	−0.199	*a*	0.497
Li−N	*a*	0.108	*a*	0.735
Li−O	*a*	0.035	*a*	0.743
Li−F	*a*	0.194	*a*	0.651
Li−S	*a*	−0.043	*a*	0.550
Li−Cl	*a*	−0.122	*a*	0.630
C−N	−0.027	−0.003	0.049	0.064
C−O	−0.185	−0.231	−0.063	−0.089
C−F	−0.126	−0.161	−0.046	−0.062
C−Si		0.073		0.137
C−P	−0.101	−0.051	0.010	0.057
C−S	0.008	−0.032	0.209	0.116
C−Cl	−0.017	−0.031	0.073	0.068
C−Br	0.019	0.005	0.034	0.024
N−O	−0.100	−0.059	−0.091	−0.074
N−P		−0.075		0.003
O−Si		0.135		0.254
O−P	0.025	0.084	0.174	0.263
O−S	0.075	0.087	0.185	0.197
F−Si		0.065		0.223
F−P	0.095	0.075	0.320	0.249
Si−Cl		−0.058		−0.074
P−S	0.045	0.222	−0.020	−0.001
P−Cl		0.058		−0.107

[a] Li is not included in the CM2 model.

4.3. Parameters for an Arbitrary Percentage of Hartree−Fock Exchange. As mentioned in the Introduction, different percentages of HF exchange are optimal for various applications. It is therefore useful to have a function that yields the parameters for *any* particular percentage of HF exchange. After parametrization of the CM3 model for all five percentages of HF exchange, we observed that in every case the parameters were either linearly or quadratically dependent on X. Therefore, the optimum parameters for an arbitrary percentage X of HF exchange can be written to a good approximation as

$$P_{ZZ'} = b_{ZZ'} + \sum_{i=1}^{1\text{ or }2} X^i m_{ZZ'}^{[i]} \qquad (7)$$

where $P_{ZZ'}$ is either $C_{ZZ'}$ or $D_{ZZ'}$. The constants in eq 7, which

TABLE 19: Root-Mean Square (RMS) Errors (in debyes) for HF/MIDI!6D and HF/6-31G* for CM2 and CM3 for the Subset of the CM2 Training Set Obtained by Removing Iodine-Containing Compounds

		HF/MIDI!6D		HF/6-31G*	
type of compound	no.[a]	CM2	CM3	CM2	CM3
H, C, N, and O polar compounds					
alcohols and water	12	0.12	0.12	0.13	0.13
esters and lactones	8	0.19	0.17	0.20	0.16
aldehydes and ketones	15	0.20	0.22	0.17	0.15
carboxylic acids	9	0.36	0.38	0.31	0.30
ethers	9	0.16	0.15	0.15	0.11
amines and ammonia	13	0.22	0.40	0.15	0.24
nitriles and HCN	17	0.18	0.19	0.24	0.20
imines and N aromatics	7	0.33	0.31	0.29	0.29
amides	3	0.32	0.25	0.24	0.25
multifunctional N	7	0.18	0.27	0.18	0.19
H, C, N, and O compounds	100	0.22	0.26	0.21	0.20
fluorides	31	0.20	0.21	0.20	0.19
chlorides	22	0.19	0.22	0.14	0.14
bromides	10	0.13	0.11	0.12	0.13
silicon	6	0.21	0.11	0.09	0.12
sulfur	9	0.35	0.47	0.12	0.46
phosphorus	13	0.36	0.23	0.41	0.40
primary training set	191	0.23	0.25	0.21	0.23
secondary training set	13	0.17	0.19	0.38	0.41
primary and secondary sets	204	0.23	0.25	0.23	0.24

[a] Number of data in the training set for this row.

TABLE 20: Root-Mean Square (RMS) Errors (in debyes) for HF/MIDI!6D and HF/6-31G* for CM2 and CM3 for the CM3 Training Set

		HF/MIDI!6D		HF/6-31G*	
type of compound	no.[a]	CM2	CM3	CM2	CM3
inorganic compounds	10	0.32	0.23	0.34	0.40
alcohols, phenol	13	0.14	0.14	0.16	0.15
ethers	11	0.20	0.19	0.18	0.13
aldehydes	5	0.16	0.21	0.22	0.19
ketones	11	0.25	0.27	0.18	0.18
carboxylic acids	9	0.36	0.38	0.31	0.30
esters	6	0.13	0.14	0.21	0.23
other C, H, and O compounds	12	0.29	0.27	0.27	0.24
aliphatic amines, aniline	13	0.22	0.40	0.15	0.23
aromatic nitrogen heterocycles	11	0.45	0.29	0.34	0.28
nitriles	12	0.20	0.20	0.23	0.23
imines	6	0.39	0.54	0.35	0.41
other C, H, and N compounds	14	0.22	0.21	0.24	0.17
amides and phenylurea	17	0.32	0.27	0.22	0.23
nitrohydrocarbons	5	0.16	0.37	0.17	0.30
bifunctional H, C, N, and O compounds	12	0.39	0.29	0.31	0.27
H, C, N, and O polar compounds[c]	163	0.28	0.28	0.25	0.24
fluorine-containing compounds	39	0.20	0.21	0.21	0.20
chlorine-containing compounds	33	0.18	0.16	0.12	0.12
bromine-containing compounds	14	0.14	0.12	0.12	0.12
halogenated bifunctional compounds	23	0.24	0.28	0.21	0.22
thiols	8	0.10	0.15	0.13	0.20
sulfides, disulfides	9	0.34	0.22	0.20	0.32
other sulfur-containing compounds	23	0.48	0.49	0.57	0.64
phosphorus	10	0.50	0.21	0.27	0.51
multifunctional phosphorus	13	0.59	0.22	0.92	0.40
compounds with S and P	7	0.60	0.14	0.78	0.35
C, H, and Si	9	0.17	0.10	0.07	0.11
C, H, O, and Si	9	0.51	0.22	0.63	0.22
C, H, Si, and halogen	18	0.78	0.16	0.84	0.13
all polar compounds excluding lithium	382	0.35	0.26	0.38	0.28
lithium compounds	16	*b*	0.59	*b*	0.27
all polar compounds	398	*b*	0.28	*b*	0.28

[a] Number of data in the training set for this row. [b] Lithium is not included in the CM2 model. [c] Six of the 10 inorganic compounds contain at most H, C, N, and O.

TABLE 21: Root-Mean Square (RMS) Errors (in debyes) for the CM2 Model Based on HF/6-31+G* Using the CM3 Model Based on mPW1PW91/6-31+G**

		CM2	CM3
type of compound	no.[a]	HF/6-31+G*	mPW1PW91/6-31+G**
inorganic compounds	10	0.45	0.36
alcohols, phenol	13	0.35	0.24
ethers	11	0.51	0.23
aldehydes	5	0.35	0.30
ketones	11	0.34	0.22
carboxylic acids	9	0.43	0.27
esters	6	0.55	0.27
other C, H, and O compounds	12	0.34	0.25
aliphatic amines, aniline	13	0.33	0.27
aromatic nitrogen heterocycles	11	0.44	0.39
nitriles	12	0.35	0.26
imines	6	0.52	0.40
other C, H, N compounds	14	0.20	0.24
amides and phenylurea	17	0.36	0.36
nitrohydrocarbons	5	0.51	0.24
bifunctional H, C, N, and O compounds	12	0.30	0.27
H, C, N, and O polar compounds[c]	163	0.39	0.28
fluorine-containing compounds	39	0.63	0.38
chlorine-containing compounds	33	0.25	0.27
bromine-containing compounds	14	0.35	0.27
halogenated bifunctional compounds	23	0.30	0.28
thiols	8	0.21	0.15
sulfides, disulfides	9	0.35	0.24
other sulfur-containing compounds	23	0.71	0.63
phosphorus	10	0.63	0.34
multifunctional phosphorus	13	1.32	0.27
compounds with S and P	7	1.42	0.38
C, H, and Si	9	0.25	0.21
C, H, O, and Si	9	1.25	0.36
C, H, Si, and halogen	18	1.07	0.65
all polar compounds excluding lithium	382	0.60	0.35
lithium compounds	16	[b]	0.41
all polar compounds	398	[b]	0.35

[a] Number of data in the training set for this row. [b] Lithium is not included in the CM2 model. [c] Six of the 10 inorganic compounds contain at most H, C, N, O.

are determined through a linear or quadratic regression, are listed in Tables 8−10, S-2, and S-3. Note that an entry of 0.0000 in the column labeled $m_{ZZ'}^{[2]}$ implies that a linear regression was used for that particular parameter, whereas a nonzero value implies that a quadratic regression was used. To test the applicability of this method, we calculated CM3 parameters using eq 7 and compared them to the parameters determined directly. For this purpose, a sixth percentage of HF exchange was chosen, 51.7, which is midway between of 42.8 and 60.6. Table 16 shows the parameters determined by both methods. The largest absolute deviation between the two sets of parameters was 0.001, which occurred for the D_{LiC} parameter. The mean absolute deviation between the two sets of parameters is 0.0005. The difference in parameters leads to only a small change in the dipole moments and in the overall error. This is shown in Table 17.

4.4. CM2 Compared to CM3. There are several changes from the CM2 model: the training set, the choice of parameters, the order of determining the parameters, the molecular geometries of molecules in the parametrization set, the choices of electronic structure theory, and the basis sets. To ascertain the relative performance of the CM2 and CM3 methods, we developed CM3 parameters for two combinations of electronic structure theory and basis set, HF/MIDI!6D and HF/6-31G*, both of which were also developed for the CM2 model. The CM2 and CM3 parameters for these two methods are shown in Table 18. It is worth noting that using 100% HF exchange does not correspond to Hartree−Fock calculations because Hartree−Fock does not include the correlation functional. The parameters

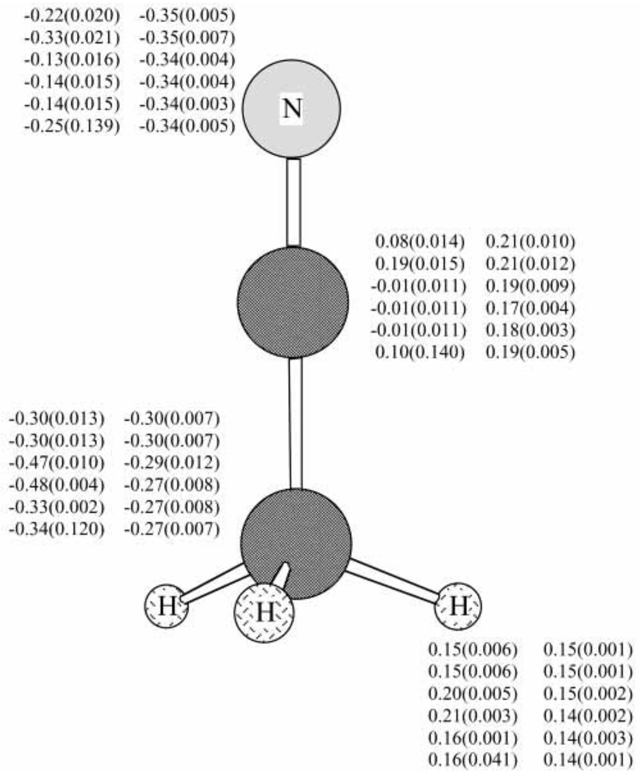

Figure 1. Löwdin and CM3 partial atomic charges in acetonitrile. (Note that charges calculated with the 6-31+G* and 6-31+G** basis sets are RLPA charges instead of Löwdin charges.) In this figure, all hydrogen atoms are equivalent, so the partial atomic charges are only shown for one of them. The left-hand column of charges for each atom are Löwdin charges, and the right-hand column of charges are CM3 charges. Row one is for MPW*X*/MIDI!, row two for MPW*X*/MIDI!6D, row three for MPW*X*/6-31G*, row four for MPW*X*/6-31+G*, row five for MPW*X*/6-31+G**, and row six for HF. In these rows, the value given first is the average charge averaged over five values of *X*, namely, 0, 25, 42.8, 60.6, and 99.9, and the value in parentheses is the standard deviation over these five values of *X*. The sixth row gives the average and standard deviation over HF/MIDI!6D and HF/6-31G*.

show, as in the hybrid DFT results, that the charge model corrections for these two HF CM3 models are typically quite small.

The performance of the CM3 model, in terms of RMS error, for compounds in the CM2 training set is given in Table 19. Over the full set, the RMS errors for the two models are equivalent for HF/6-31G*; however, *most* of the classes of compounds show increased accuracy when going from CM2 to CM3. The CM3 model performs noticeably worse than CM2 for two classes of compounds in the CM2 training set, amines and sulfur compounds. The RMS errors for all of the compounds in the CM3 training set are given in Table 20. This table shows that the worse performance for amines is primarily a tradeoff that is correlated with improved performance for nitrogen heterocycles. The overall error for H, C, N, and O compounds is almost unchanged. However, CM3 appears to be much better than CM2 for Si and P compounds.

Table 21 compares the CM2 HF/6-31+G* parametrization to one of the methods parametrized here, namely, the CM3 model based on mPW1PW91/6-31+G**. The RMS error over compounds that contain at most H, C, N, and O are nearly the same for both methods. However, the RMS errors over compounds that contain F and Br for the CM3 model based on mPW1PW91/6-31+G** are smaller than the corresponding RMS errors for the CM2 HF/6-31+G* model. This CM3 model additionally shows significant improvements over the CM2 HF/6-31+G* model for molecules containing Si, P, and S.

TABLE 22: Mean-Unsigned Error (MUE, and in debyes) over the 398 Data

	X				
	0	25	42.8	60.6	99.9
MPW*X*/MIDI!5D	0.20	0.19	0.19	0.19	0.20
MPW*X*/MIDI!6D	0.20	0.20	0.19	0.19	0.20
MPW*X*/6-31G*	0.19	0.19	0.19	0.19	0.20
MPW*X*/6-31+G*	0.28	0.27	0.27	0.27	0.28
MPW*X*/6-31+G**	0.26	0.26	0.27	0.27	0.28
HF/MIDI!6D					0.20[a]
HF/6-31G*					0.20[a]

[a] $X = 100\%$, no correlation functionals.

4.5. Partial Atomic Charges. To show that CM3 partial atomic charges remain reasonably constant for a wide range of percentages of HF exchange and for a variety of basis sets, we present the average Löwdin and CM3 partial atomic charges for 10 representative molecules, namely, acetonitrile, methanol, formaldehyde, α-d-glucose,[73] methylamine, formamide, chloromethane, methanethiol, and methylsilane in Figure 1 and S-1 to S-9 of the Supporting Information. Note that for the 6-31+G* and 6-31+G** basis sets we use RLPA charges instead of Löwdin charges. Figures 1 and S-1 to S-8 give, for each symmetry-unique atom in the molecules described above, the average of the Löwdin (or RLPA) and CM3 partial atomic charges calculated by five percentages of Hartree−Fock exchange, namely, 0, 25, 42.8, 60.6, and 99.9, with MPW*X* for each of the five basis sets used in this paper. Also given is the average of the Löwdin and CM3 partial atomic charges calculated with the two Hartree−Fock methods used in this paper, HF/MIDI!6D and HF/6-31G*. The standard deviation of each average partial atomic charge, which indicates how consistent the charge distribution is for different percentages of Hartree−Fock exchange, is also included. The standard deviations of the CM3 partial atomic charges are smaller than the standard deviations of the Löwdin partial atomic charges for the majority of the atoms in this representative set of molecules. Furthermore, and even more significantly, the CM3 results do not vary as much from row to row as the Löwdin charges do, although the variations from row to row are bigger than the standard deviations within the rows.

4.6. Mean Unsigned Errors. In general, for fits where the errors are not expected to be distributed normally, the mean unsigned error (mean absolute value of the error) may be a more robust indicator of the quality of the fit than the root-mean-squared error.[74] Table 22 provides this quantity for all 27 parametrizations of the present paper.

5. Conclusions

We have presented a model for the prediction of atomic partial charges based on hybrid density functional theory with variable Hartree−Fock exchange for five basis sets. In addition, we have presented parametrizations for two basis sets with Hartree−Fock theory. All of these parametrizations have RMS errors that are between 0.26 and 0.40 D and mean unsigned errors in the range 0.19−0.28 D.

Acknowledgment. The authors are grateful to Timothy M. Sonbuchner for participation in the early stages of this work. This work was supported in part by the National Science Foundation under Grant Nos. CHE00-92019 and CHE-9876792 and by the U. S. Army Research Office under the Multidisciplinary University Research Initiative (MURI) program, Grant No. DAAD19-02-0176.

Supporting Information Available: Additional tables such as Tables 8−13 but for the MIDI!6D and 6-31+G* basis sets, and Löwdin (or RLPA) and CM3 partial atomic charges for eight of the nine compounds discussed in section 4.5, i.e., methanol, formaldehyde, α-d-glucose, formamide, methylamine, chloromethane, methanethiol, and methylsilane (14 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Coppens, P. *Annu. Rev. Phys. Chem.* **1992**, *43*, 663.

(2) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aid. Mol. Des.* **1995**, *9*, 87.

(3) Mulliken, R. S. *J. Chem. Phys.* **1935**, *3*, 564.

(4) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(5) Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428.

(6) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365.

(7) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.

(8) Bader, R. F. *Acc. Chem. Res.* **1985**, *18*, 9.

(9) Momany, F. A. *J. Phys. Chem.* **1978**, *86*, 592.

(10) Kollman, P. A. *J. Am. Chem. Soc.* **1977**, *99*, 4875.

(11) Kollman, P. A. *J. Am. Chem. Soc.* **1978**, *100*, 2974.

(12) Smit, P. H.; Derissen, J. L.; Van Duijneveldt, P. B. *Mol. Phys.* **1988**, *37*, 521.

(13) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304.

(14) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(15) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.

(16) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(17) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(18) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(19) Merz, K. M. *J. Comput. Chem.* **1992**, *13*, 749.

(20) Montagnani, R.; Tomasi, J. *J. Mol. Struct. (THEOCHEM)* **1993**, *279*, 131.

(21) Hall, D.; Williams, D. E. *Acta Crystallogr.* **1975**, *A31*, 56.

(22) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.

(23) Li, J.; Williams, B.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 724.

(24) Li, J.; Xing, J.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *111*, 885.

(25) Zerner, M. C. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1991; Vol. 2, p 313.

(26) Simons, J.; Nichols, J. *Quantum Mechanics in Chemistry*; Oxford University Press: New York, 1997.

(27) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

(28) Becke, A. *J. Chem. Phys.* **1993**, *98*, 1372.

(29) Becke, A. D. *J. Chem. Phys.* **2000**, *112*, 4020.

(30) Abu-Awwad, F.; Politzer, P. *J. Comput. Chem.* **2000**, *21*, 227.

(31) (a) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811. (b) Kormos, B. K.; Cramer, C. J. *J. Phys. Org. Chem.* **2002**, *15*, 71.

(32) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(33) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(34) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.

(35) Perdew, J. P. In *Electronic Structure of Solids '91*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; p 11.

(36) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 2936.

(37) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. *PhysChemComm* **2002**, *5*, 117.

(38) Mayer, I. *Chem. Phys. Lett.* **1983**, *97*, 270.

(39) Mayer, I. *Chem. Phys. Lett.* **1985**, *117*, 396.

(40) Mayer, I. *Int. J. Quantum Chem.* **1986**, *29*, 73.

(41) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281.

(42) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192.

(43) Thompson, J. D.; Winget, P.; Truhlar, D. G. *PhysChemComm* **2001**, *4/16*, 1.

(44) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.

(45) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(46) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972**, *66*, 217.

(47) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.

(48) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.

(49) Binning, R. C., Jr.; Curtiss, L. A. *J. Comput. Chem.* **1990**, *11*, 1206.

(50) *CRC Handbook of Chemistry and Physics*, 81st ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 2000.

(51) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. Manuscript in preparation.

(52) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407.

(53) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703.

(54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.11; Gaussian, Inc.: Pittsburgh, PA, 1998.

(55) Xidos, J. D.; Li, J.; Thompson, J. D.; Hawkins, G. D.; Winget, P.; Zhu, T. H.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. MN-GSM, *version 1.8*; University of Minnesota: Minneapolis, 2002.

(56) *NAG Fortran 90 Library*, 4 ed.; The Numerical Algorithms Group, Inc.: Oxford, 2000.

(57) McClellan, A. L. *Tables of Experimental Dipole Moments*; W. H. Freeman: San Fransisco, CA, 1963.

(58) Nelson, R. D.; Lide, D. R.; Maryott, A. A. *Natl. Stand., Ref. Data Ser.*; United States National Bureau of Standards NSRDS−NBS 10: Washington, DC, 1967.

(59) Stark, B. In *Molecular Constants from Microwave Spectroscopy*; Landolt-Börnstein, New Series, Group II; Hellwege, K.-H., Hellwege, A. M., Eds.; Springer-Verlag: Berlin, 1967; Vol. 4, p 136.

(60) Stark, B. In *Molecular Constants from Microwave Spectroscopy*; Landolt-Börnstein, New Series, Group II; Hellwege, K.-H., Hellwege, A. M., Eds.; Springer-Verlag: Berlin, 1982; Vol. 14a, p 261.

(61) Demaison, J.; Hünter, W.; Stark, B.; Buck, I.; Tischer, R.; Winnewisser, M. In *Molecular Constants*; Landolt-Börnstein, New Series, Group II; Hellwege, K.-H., Hellwege, A. M., Eds.; Springer-Verlag: Berlin, 1974; Vol. 6, p 261.

(62) Abraham, R. J.; Grant, G. H. *J. Comput. Chem.* **1988**, *9*, 244.

(63) Gerber, P. R. *J. Comput.-Aid. Mol. Des.* **1988**, *12*, 37.

(64) Gillies, J. Z.; Cotter, E.; Gilles, C. W.; Warner, H. E.; Block, E. *J. Phys. Chem. A* **1999**, *103*, 4948.

(65) Mayer, I. *Chem. Phys. Lett.* **1984**, *110*, 44.

(66) Cioslowski, J. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III., Eds.; Wiley: Chickester, U.K., 1998; p 892.

(67) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.

(68) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, *61*, 8720.

(69) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys*. **1998**, *109*, 9117; errata: **1999**, *111*, 5624; **2000**, *113*, 3930.

(70) Li. J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293.

(71) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 63.

(72) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys*. **1999**, *110*, 5503.

(73) Barrows, S. E.; Storer, J. W.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *J. Comput. Chem*. **1998**, *19*, 1111.

(74) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1992; p 605.